

Introduction

The tephritid fly, *Ceratitis cosyra* (Walker, 1849), is possibly the most important indigenous pest of mango throughout sub-Saharan Africa. It is estimated that *C. cosyra* can reduce the mango crop yield between 20 and 30%, and the damage this pest causes affects the quality and market value of the fruit at both local and international markets (Lux et al. 2003). Although it has been partially displaced by the invasive *Bactrocera dorsalis* (Hendel, 1912) in recent years (Ekesi et al. 2009), it still has serious impact on the mango production, especially for early and mid-season cultivars in western Africa (Vayssières et al. 2009). The range of host records for *C. cosyra* is relatively large (but not as large as for other congeneric fruit flies such as *C. rosa* Karsch, 1887 or *C. capitata* (Wiedemann, 1824), see Copeland et al. 2006 for details) and includes guava as well as a variety of hosts from Anacardiaceae, Annonaceae, Apocynaceae and Rubiaceae among others (De Meyer 1998).

In Kenya, the mango fruit fly can be found in both lowlands and highlands at altitudes between 20 and 2100 m, while Geurts et al. (2012) observed predominance at lower altitudes in Tanzania. On the Kenyan coast, *C. cosyra* uses wild fruits, primarily the marula tree, *Sclerocarya birrea* (A. Rich.) Hochst., as an alternative host when mango is not available (Copeland et al. 2006). Similarly, in Tanzania *C. cosyra* shifts to soursop (*Annona muricata* Linnaeus) out of the mango season (Mwatawala et al. 2009). The distribution of *C. cosyra* in southern Africa is limited to the subtropical regions and its occurrence in this region coincides with the known distribution range of *S. birrea* (De Meyer 2001). *Ceratitis cosyra* is commonly intercepted in Europe (Li et al. 2009 and references therein) where the establishment of adventive populations raises serious concerns. Based on the observed altitudinal records, Copeland et al. (2006) suggested that *C. cosyra* may be pre-adapted to survive in the same subtropical and Mediterranean climatic areas as the cosmopolitan *C. capitata*, thus representing a potential risk of invasion and establishment of this pest to Europe and the US mainland (but see Grout and Stoltz 2007 for a less pessimistic view on its invasion potential). Similarly, Li et al. (2009) listed North, Central and South American and Middle Eastern, Asian and Australian countries as potential suitable areas for the establishment of adventive *C. cosyra* populations.

Barr et al. (2012) investigated the utility of DNA barcoding for molecular identification of several tephritid pests, including *C. cosyra*. They suggested that the mango fruit fly might include cryptic species. In fact, in their study, *C. cosyra* was represented by a larger haplotype group with vouchers from Mali (two sampling locations) and Kenya (two sampling locations) and by two *C. cosyra* outliers sampled at the coast of southern Kenya (Shimba Hills). Surprisingly, these two individuals, sharing the same COI haplotype, were separated from the main haplotype group by 52 mutational steps. Also on morphological grounds, *C. cosyra* has been the subject of confusion. Various taxa (now considered synonyms) have been described as separate species or varieties based on differences in cephalic and leg chaetotaxy and mesonotal patterns (see De Meyer 1998 for detailed discussion). The occurrence of cryptic species in *C. cosyra*

would pose relevant issues with respect to pest management, ecological modelling and estimation of its invasion potential. In this study, the population structure of *C. cosyra* was inferred across the species distributional range as a first step towards exploring its cryptic diversity.

Methods

A total of 348 specimens of *C. cosyra* from 13 populations ($13 < n < 32$) were collected in Africa from 2000 to 2012 (Table 1, see also supplementary file SF 1: Map of sampling locations). DNA was extracted from ethanol preserved adults by the DNeasy Blood and Tissue Kit (Qiagen) as per the manufacturer's instructions. Individual flies were genotyped at 16 polymorphic microsatellite loci developed by Delatte et al. (2014): Co1350, Co1444, Co2J, Co486, Co633, Co806, CoD4, CoES, CoKW, CoOI, CoP7, CoQT, CoRTA, CoWU, CoZ29, CoZW (see Delatte et al. (2014) for primer sequences and laboratory procedures). Electrophoretic analyses were conducted on an automated ABI Prism 3100 Genetic Analyzer (Applied Biosystem) with individuals declared non-amplifiable at a locus after two independent amplification failures. The genotypes of the 348 individual insects were analysed by the ADEGENET 1.4-2 package of the R statistical software (Jombart 2008) to ascertain the genetic variability and differentiation, among the *C. cosyra* populations, including number of alleles per locus (N_{all}), observed and expected heterozygosity (H_{obs} , H_{exp}) and deviations from the Hardy-Weinberg equilibrium (HWE). The function *genotype_curve* of the R package POPR (Kamvar et al. 2014) was used to calculate a genotype accumulation curve (this function randomly sample loci without replacement and count for the number of multilocus genotypes). Linkage disequilibrium was tested for each population across each pair of loci using the log likelihood ratio statistic implemented in GENEPOP 4.3 (Rousset 2008) and assessing significance through Markov-chain randomizations based on 1000 dememorizations, 100 batches, and 5000 iterations per batch. FreeNA 1.0 (Chapuis and Estoup 2007) was used to estimate null allele frequencies (per locus and population) according to Dempster et al. (1977). Probability values of repeated tests were corrected for Type I errors using the False Discovery Rate (FDR) procedure (Benjamini and Hochberg 1995). Isolation by distance (IBD) was verified in ADEGENET through Mantel test between Edwards' genetic distances and Euclidean geographic distances (1000 permutations). Principal Component Analysis (PCA) was used to ordinate specimens in multivariate space. Prior to PCA, the SCALEGEN function of ADEGENET was used to centre the data and replace missing genotypes with mean allele frequencies. Specimens from different populations were then ordinated by maximising between-group variances through Discriminant Analysis of Principal Components (DAPC). The number of Principal Components (PCs) retained in DAPC was optimised using XVALDAPC function of ADEGENET (Jombart et al. 2010).

STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to calculate individual admixture coefficients (Q) across individuals and populations. STRUCTURE analyses

were based on the admixture model (individuals were allowed to have mixed ancestries from different clusters) with correlated allele frequencies (allele frequencies in different clusters were likely to be similar due to migration or shared ancestry) and the parameter of the Dirichlet distribution of allelic frequencies (λ) separately inferred for each population. We used STRUCTURE HARVESTER 0.6.94 (Earl and vonHoldt 2012) to infer the optimal number of clusters (K) using the Evanno et al. (2005) parameters. Since this method only detects the uppermost level of population structure when different hierarchical levels exist, we further investigated the genetic substructuring of our dataset by following the sequential clustering method described in Coulon et al. (2008). For this purpose, replicated STRUCTURE runs were performed by (a) dividing the main dataset in subsets of data including individuals assigned to the same cluster, (b) recalculating the optimal K value (Evanno et al. 2005) of each subset of data and (c) repeating the STRUCTURE analyses of each subset of data. We set $Q = 0.7$ as an arbitrary threshold for cluster assignment and individuals not reaching the threshold were discarded from further replicated runs. For each value of K, five iterations were run for 3 million generations (with 1.5 million generations as burn-in) and the posterior estimates of cluster memberships of the 3 runs with the highest estimated log probability of the data were summarized in CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) and visualized in DISTRUCT 1.1 (Rosenberg 2004).

Results

The amount of scored multilocus genotypes reached a plateau after 5–7 sampled loci, indicating that the genetic variability of *C. cosyra* was adequately sampled by the 16 microsatellites markers used (see supplementary file SF 2: Genotype accumulation curve). The total scored number of alleles (N_{all}) ranged from 67 (in the Burkina Faso population) to 124 (Mozambique), with an average proportion of missing data per population ranging from 2.2% (SE = 1.5%, Mali) to 30.9% (SE = 9.0%, Ivory Coast). H_{obs} ranged from 0.372 (Mali) to 0.504 (South Africa), while H_{exp} from 0.432 (Mali) to 0.637 (Tanzania) (Table 1). Pearson's Chi-squared test showed significant HWE deviations in 84 of 208 locus-population combinations, corresponding to 40.4% of observations (see supplementary files SF 3: Pearson's Chi-squared test for Hardy-Weinberg equilibrium and SF 4: Observed and expected heterozygosity). These HWE deviations are compatible with the occurrence of genetic substructuring within populations (Walhund effect) as described below. The average estimated proportion of null alleles was 8.1% (SE = 0.6%) (see supplementary file SF 5: Estimated null allele proportions) and linkage disequilibrium was observed in 16.7% of pairwise tests (see supplementary file SF 6: Linkage disequilibrium). Mantel test did not evidence significant correlation between individual geographic and genetic distances ($p > 0.05$, see supplementary file SF 7: Mantel test).

PCA was based on 28 PC axes that accounted for 70.6% of cumulative inertia. The first two PCs (Figure 1) represented a relatively low amount of variation (22.3%)

Table 1. Population locations and genetic variability. Sampling locations, geographic coordinates (decimal degrees) and summary of genetic variability in 13 populations of *C. cosyra* (see Figure 1) genotyped at 16 microsatellite loci. N: number of individuals per population, N_{all}: total number of alleles, H_{obs}: observed heterozygosity, H_{exp}: expected heterozygosity, null: mean null allele frequency based on Dempster et al. (1977). Standard deviations in parentheses.

| | Locality | | Latitude | Longitude | N | N _{all} | H _{obs} | H _{exp} | null |
|----|--------------|----------------|----------|-----------|----|------------------|------------------|------------------|---------------|
| 1 | Burkina Faso | (interception) | | | 29 | 67 | 0.412 (0.319) | 0.484 (0.289) | 0.060 (0.094) |
| 2 | Burundi | Isabu | -3.394 | 29.361 | 32 | 79 | 0.391 (0.312) | 0.478 (0.304) | 0.070 (0.090) |
| 3 | Ethiopia | Badano | 9.317 | 41.217 | 13 | 74 | 0.430 (0.313) | 0.475 (0.279) | 0.065 (0.073) |
| 4 | Ivory Coast | Korhogo | 9.450 | -5.633 | 18 | 100 | 0.417 (0.327) | 0.449 (0.366) | 0.036 (0.059) |
| 5 | Kenya | Nairobi | -1.283 | 36.817 | 32 | 122 | 0.415 (0.276) | 0.557 (0.266) | 0.109 (0.062) |
| 6 | Malawi | Zomba | -15.383 | 35.333 | 29 | 118 | 0.481 (0.266) | 0.601 (0.266) | 0.082 (0.090) |
| 7 | Mali | (interception) | | | 29 | 69 | 0.372 (0.308) | 0.432 (0.331) | 0.050 (0.073) |
| 8 | Mozambique | Cuamba | -14.816 | 36.535 | 32 | 124 | 0.478 (0.276) | 0.581 (0.255) | 0.076 (0.092) |
| 9 | Nigeria | Sokoto | 13.051 | 5.231 | 26 | 93 | 0.421 (0.292) | 0.482 (0.283) | 0.059 (0.074) |
| 10 | South Africa | Constantia | -23.644 | 30.679 | 22 | 110 | 0.504 (0.284) | 0.570 (0.278) | 0.061 (0.075) |
| 11 | Senegal | Sané | 12.750 | -15.500 | 28 | 114 | 0.393 (0.262) | 0.559 (0.259) | 0.112 (0.112) |
| 12 | Sudan | Singa | 13.150 | 33.850 | 32 | 113 | 0.376 (0.287) | 0.532 (0.279) | 0.123 (0.091) |
| 13 | Tanzania | Mzinga | -6.883 | 37.617 | 26 | 101 | 0.410 (0.255) | 0.637 (0.186) | 0.146 (0.113) |

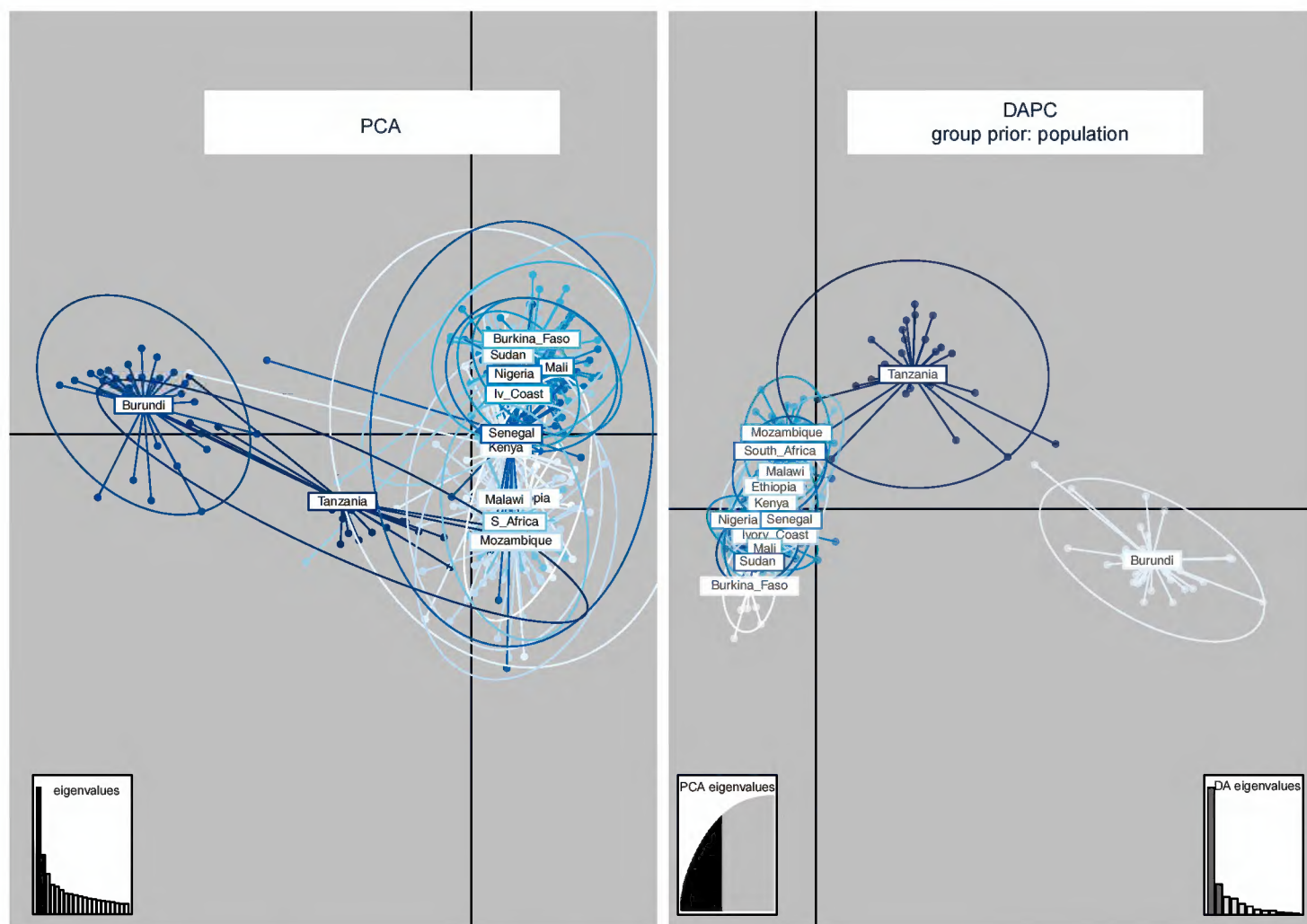


Figure 1. Unconstrained and constrained ordination. Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC) of 348 *C. cosyra* microsatellite genotypes. Specimen groups are labelled inside their 95% inertia ellipses and genotypes are connected to the corresponding group centroids.

and didn't allow proper resolution of populations (see 95% confidence ellipses). DAPC considered the populations as a priori defined groups and was based on 90 PCs. Stressing the ordination of points by using a prior group only allowed a better resolution of the population from Burundi and, possibly, of that from Tanzania. The latter was clearly separated from the other populations only when excluding 4 STRUCTURE outliers from DAPC (see below). Confidence ellipses of all other populations were largely overlapping (Figure 1).

The STRUCTURE analysis of the entire dataset ($n=345$, run 0, Figure 2) showed ΔK values (Evanno et al. 2005) peaking at $K=2$ indicating that the main hierarchical level of the population structure is based on 2 genotype groups. A first and smaller group (then separately analysed in run 1) included all specimens from Burundi ($n=32$) and 22 out of the 26 individuals from Tanzania. A second and larger group (subsequently analysed in run 2) included 98.6% of genotypes from all other populations combined ($n=290$), including the 4 outliers from Tanzania (see supplementary file SF 8: STRUCTURE sequential assignments). Run 1 ($K=2$) resolved all specimens from Burundi in one group (that also included one outlier from Kenya and one from Senegal), while specimens from Tanzania were partially assigned to the Burundi group (5 specimens) and, for a larger part, to a second group (17 specimens) together with

sequential Bayesian assignment

following Coulon et al. (2008)

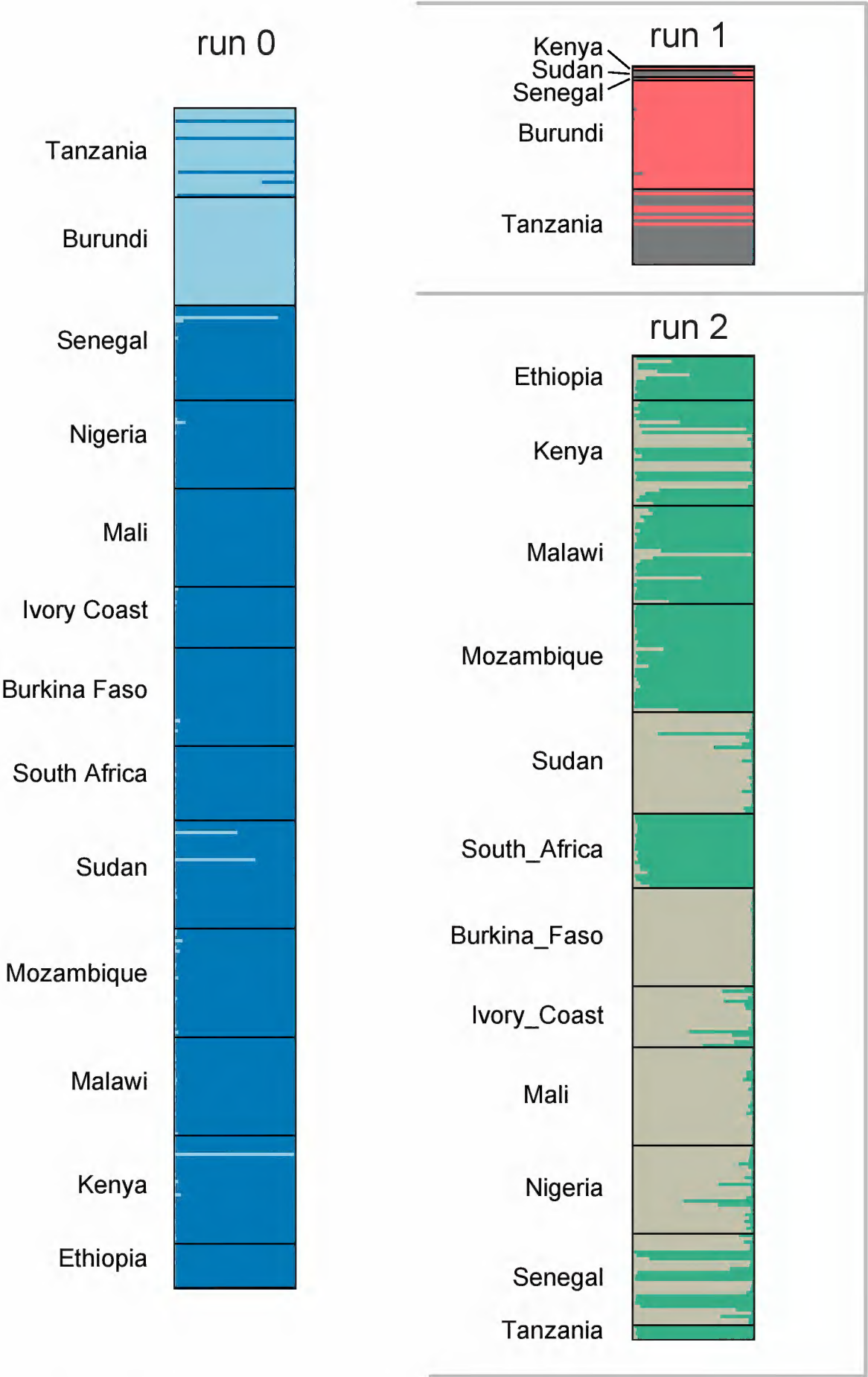


Figure 2. Individual Bayesian assignments. STRUCTURE sequential individual assignments of 348 specimens of *C. cosyra* from 13 African countries.

2 outliers from Sudan. Run 2 ($K=2$) resolved specimens from South Africa (100%), Mozambique (96.9%), Malawi (93.1) and Ethiopia (84.6%) in a first group and specimens from Burkina Faso (100%), Mali (100%), Nigeria (96.2%), Sudan (87.5%) and Ivory Coast (88.9%) in a second group. Populations from Senegal and Kenya included specimens that were assigned in part to the first group (35.7% and 62.5%, respectively) and in part to the second (60.7% and 31.3%).

Discussion

The morphospecies *C. cosyra* includes two groups of genetically well-differentiated individuals. The vast majority of vouchers from Burundi and Tanzania (plus a few outliers from other African countries) belong to the first of these two groups, all other specimens to the second. Specimens from the two groups were also found sympatrically in a number of populations from Kenya, Senegal, Sudan, and Tanzania. Interestingly, the two *C. cosyra* outliers described by Barr et al. (2012) were also sampled from the Kenyan coast. The fact that our Kenyan population included specimens from the two clusters suggests that Barr et al. might have genotyped representatives of both types. If this would be confirmed, the two genotypes resolved through microsatellite markers might also show marked differentiation in the cytochrome oxidase I gene barcode region so that they might be more easily diagnosed through DNA barcoding rather than through microsatellite genotyping.

Further studies are needed to verify if specimens from the two genotypic clusters are also morphologically, ecologically and / or physiologically different and to which extent the two groups are connected by gene flow. A preliminary screening of thorax patterns of the samples used in this study (6 characters scored, data not shown) did not reveal straightforward morphological differences between groups. Wing morphometrics (Van Cann et al. 2015) might provide a more suitable tool to further investigate morphological cryptic variation in *C. cosyra*.

A wide variety of hosts have been described for *C. cosyra*, including Annonaceae (such as the introduced soursop) or Anacardiaceae (including the indigenous marula or the introduced mango). An intriguing hypothesis is that the two different *C. cosyra* types might also have different host preferences, similarly to what has been observed by McPherson et al. (1988) for *Rhagoletis pomonella* (Walsh, 1867). A separate host range characterization for the two types might provide useful information and help to understand if the observed genetic split has a recent evolutionary history (possibly related to the introduction of novel hosts in Africa) or a deeper evolutionary origin.

The sequential Bayesian assignment of genotypes also helped to disentangle the effects of cryptic speciation and of population structure within each of the genotypic groups. Specimens from Burundi and Tanzania are, to a less extent, genetically divergent, and among samples from the other African countries, two groups can be further resolved. In the latter case, specimens could be roughly subdivided between Western African samples (including Burkina Faso, Ivory Coast, Mali and Nigeria) and Eastern / Southern African samples (including Ethiopia, Tanzania, Malawi, Mozambique, South Africa) with

the notable exception of Sudan (which is genetically closer to the West African samples) and of Kenya and Senegal (that included a mix of individuals from both groups). Morphological differences were considered by De Meyer (1998) to represent a plausible (but incomplete) geographical split between western and eastern Africa, and this is also only partially corroborated by the genotypic clustering of this study. On the other hand the preliminary morphological screening of thoracic patterns does not provide any support for this division. Most importantly, it is not clear to what extent the mixed patterns of Kenya and Senegal and Tanzania can be related to historical evolutionary processes or to more recent events involving fruit trade and transport (Malacrida et al. 2007).

Conclusions

Marked and sympatric genetic splits are compatible with the occurrence of presumptive cryptic species, within *C. cosyra*. Additional integrative taxonomy, possibly combining morphological, ecological and physiological data (e.g., see Schutze et al. 2015) is now required to further support this model.

Acknowledgements

We wish to thank Valérie Balmès and all colleagues and collectors who provided specimens for this study. This research was part of Research Project “Development of Molecular and Morphological Diagnostics for Taxonomic and Population Genetic Characterization of African Fruit Fly Pests (16032/R0)” supported by the International Atomic Energy Agency (IAEA). We would like to thank Allen Norrbom for insightful comments on an earlier version of this ms.

References

- Barr NB, Islam MS, De Meyer M, McPherson BA (2012) Molecular Identification of *Ceratitis capitata* (Diptera: Tephritidae) Using DNA Sequences of the COI Barcode Region. *Annals of the Entomological Society of America* 105: 339–350. doi: 10.1603/an11100
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society (Series B)* 57: 289–300.
- Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631. doi: 10.1093/molbev/msl191
- Copeland R, Wharton R, Luke Q, De Meyer M, Lux S, Zenz N, Machera P, Okumu M (2006) Geographic distribution, host fruit, and parasitoids of African fruit fly pests *Ceratitis anonae*, *Ceratitis cosyra*, *Ceratitis fasciventris*, and *Ceratitis rosa* (Diptera : Tephritidae) in Kenya. *Annals of the Entomological Society of America* 99: 261–278. doi: 10.1603/0013-8746(2006)099[0261:gdhfap]2.0.co;2

- Coulon A, Fitzpatrick JW, Bowman R, Stith BM, Makarewich CA, Stenzler LM, Lovette IJ (2008) Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma caerulescens*). *Molecular Ecology* 17: 1685–1701. doi: 10.1111/j.1365-294X.2008.03705.x
- De Meyer M (1998) Revision of the subgenus *Ceratitidis* (*Ceratalaspis*) Hancock (Diptera: Tephritidae). *Bulletin of Entomological Research* 88: 257–290. doi: 10.1017/S0007485300025888
- De Meyer M (2001) Distribution patterns and host-plant relationships within the genus *Ceratitidis* MacLeay (Diptera: Tephritidae) in Africa. *Cinebasia* 17: 219–228.
- Delatte H, Virgilio M, Simiand C, Quilici S, Nzogela YB, Meyer M (2014) Isolation and characterisation of sixteen microsatellite markers amplifying an African agricultural pest, *Ceratitidis cosyra* (Walker) (Diptera: Tephritidae). *Conservation Genetics Resources* 6: 9–11. doi: 10.1007/s12686-013-0026-2
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society (Series B)* 39: 1–38.
- Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361. doi: 10.1007/s12686-011-9548-7
- Ekesi S, Billah MK, Nderitu PW, Lux SA, Rwomushana I (2009) Evidence for Competitive Displacement of *Ceratitidis cosyra* by the Invasive Fruit Fly *Bactrocera invadens* (Diptera: Tephritidae) on Mango and Mechanisms Contributing to the Displacement. *Journal of Economic Entomology* 102: 981–991.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Geurts K, Mwatawala M, De Meyer M (2012) Indigenous and invasive fruit fly diversity along an altitudinal transect in Eastern Central Tanzania. *Journal of Insect Science* 12: 12. doi: 10.1673/031.012.1201
- Grout TG, Stoltz KC (2007) Developmental rates at constant temperatures of three economically important *Ceratitidis* spp. (Diptera : Tephritidae) from southern Africa. *Environmental Entomology* 36: 1310–1317. doi: 10.1603/0046-225x(2007)36[1310:dracto]2.0.co;2
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806. doi: 10.1093/bioinformatics/btm233
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405. doi: 10.1093/bioinformatics/btn129
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 1–15. doi: 10.1186/1471-2156-11-94
- Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281. doi: 10.7717/peerj.281

- Li BN, Ma J, Hu XN, Liu HJ, Zhang RJ (2009) Potential geographical distributions of the fruit flies *Ceratitis capitata*, *Ceratitis cosyra*, and *Ceratitis rosa* in China. *Journal of Economic Entomology* 102: 1781–1790. doi: 10.1603/029.102.0508
- Lux SA, Ekesi S, Dimbi S, Mohamed S, Billah M (2003) Mango infesting fruit flies in Africa. Perspectives and limitations of biological approaches to their management. In: Neuen-schwander P, Borgemeister C, Langewald J (Eds) *Biological control in integrated pest management systems in Africa*. CABI, Wallingford, 277–293.
- Malacrida A, Gomulski L, Bonizzoni M, Bertin S, Gasperi G, Guglielmino C (2007) Globalization and fruitfly invasion and expansion: the medfly paradigm. *Genetica* 131: 1–9. doi: 10.1007/s10709-006-9117-2
- McPheron BA, Smith DC, Berlocher SH (1988) Genetic differences between host races of *Rhagoletis pomonella*. *Nature* 336: 64–66. doi: 10.1038/336064a0
- Mwatawala MW, De Meyer M, Makundi RH, Maerere AP (2009) Host range and distribution of fruit-infesting pestiferous fruit flies (Diptera, Tephritidae) in selected areas of Central Tanzania. *Bulletin of Entomological Research* 99: 629–641. doi: 10.1017/S0007485309006695
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138. doi: 10.1046/j.1471-8286.2003.00566.x
- Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106. doi: 10.1111/j.1471-8286.2007.01931.x
- Schutze MK, Aketarawong N, Amornsak W, Armstrong KF, Augustinos AA, Barr N, Bo W, Bourtzis K, Boykin LM, Cáceres C, Cameron SL, Chapman TA, Chinvinijkul S, Chomič A, De Meyer M, Drosopoulou E, Englezou A, Ekesi S, Gariou-Papalexidou A, Geib SM, Hailstones D, Hasanuzzaman M, Haymer D, Hee AKW, Hendrichs J, Jessup A, Ji Q, Khamis FM, Krosch MN, Leblanc LUC, Mahmood K, Malacrida AR, Mavragani-Tsipidou P, Mwatawala M, Nishida R, Ono H, Reyes J, Rubinoff D, San Jose M, Shelly TE, Srikachar S, Tan KH, Thanaphum S, Haq I, Vijaysegaran S, Wee SL, Yesmin F, Zacharopoulou A, Clarke AR (2015) Synonymization of key pest species within the *Bactrocera dorsalis* species complex (Diptera: Tephritidae): taxonomic changes based on a review of 20 years of integrative morphological, molecular, cytogenetic, behavioural and chemoecological data. *Systematic Entomology* 40: 456–471. doi: 10.1111/syen.12113
- Van Cann J, Virgilio M, Jordaens K, De Meyer M (2015) Wing morphometrics as a possible tool for the diagnosis of the *Ceratitis fasciventris*, *C. anonae*, *C. rosa* complex (Diptera, Tephritidae). In: De Meyer M, Clarke AR, Vera MT, Hendrichs J (Eds) *Resolution of Cryptic Species Complexes of Tephritid Pests to Enhance SIT Application and Facilitate International Trade*. *ZooKeys* 540: 489–506. doi: 10.3897/zookeys.540.9724
- Vayssières JF, Korie S, Ayegnon D (2009) Correlation of fruit fly (Diptera Tephritidae) infestation of major mango cultivars in Borgou (Benin) with abiotic and biotic factors and assessment of damage. *Crop Protection* 28: 477–488. doi: 10.1016/j.cropro.2009.01.010

Map of sampling locations

Data type: species data

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Genotype accumulation curve

Data type: data analysis

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Pearson's Chi-squared test for Hardy-Weinberg equilibrium

Data type: data analysis

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

STRUCTURE sequential assignments

Data type: data analysis

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.